This listing of claims will replace all prior versions, and listings, of claims in the application:

1. - 18. **(Canceled)**

19. **(Currently Amended)** A multi-point conference device, communicatively connected to a plurality of terminals, comprising:

a medium processing unit for detecting a speaker;

a memory unit for holding an image from a terminal participating in a conference; and

an image processing unit for decoding an image of a speaker and for re-encoding the [[so]] decoded image and providing an intra frame, when the speaker is detected at speaker switching;

said image processing unit configured to transmit transmitting an the intra frame as an image frame at the time of speaker switching, when said medium processing unit detects a speaker.

20. **(Previously Presented)** A multi-point conference system comprising:

a plurality of terminals; and

the multi-point conference device, as set fourth in claim 19, the multi-point conference device connected to said plurality of terminals and transmitting/receiving image and audio to perform a conference.

21. **(Previously Presented)** The multi-point conference system as defined in claim 20, wherein said image processing unit comprises:

a decoder unit for decoding an image of a speaker held in said memory unit based on the result of speaker detection by said medium processing unit;

a reference image memory unit for holding a reference image obtained on decoding by said decoder unit the last image of a speaker held in said memory unit; and

an encoder unit for re-encoding an image obtained on decoding by said decoder unit an image received after a speaker is detected, based on a reference image held in said reference image memory unit;

wherein at least the first frame of the image of a speaker, received after a speaker is detected, is encoded as an intra frame.

22. **(Previously Presented)** The multi-point conference system as defined in claim 20, wherein said terminals and said multi-point conference device are capable of communicating with each other via a communication protocol equipped with no re-transmission procedure.

23. **(Currently Amended)** The multi-point conference device as defined in claim 19, further comprising a transmission unit configured for transmitting an intra frame re-encoded by said image processing unit as an image frame at the time of speaker switching when said medium processing unit detects a speaker.

24. **(Previously Presented)** The multi-point conference device as defined in claim 19, wherein the image processing unit comprises:

a decoder unit for decoding an image of a speaker held in said memory unit according to a speaker detection result;

a reference image memory unit for holding a reference image obtained on decoding by said decoder unit the last image of a speaker saved in said memory unit; and

an encoder unit for re-encoding an image obtained on decoding by said decoder unit an image received after a speaker is detected, based on a reference image held in said reference image memory unit; wherein

at least the first frame of the image of a speaker received after a speaker is detected is encoded as an intra frame.

25. **(Previously Presented)** The multi-point conference system as defined in claim 20, wherein the multi-point conference system connects a first network and a second network that is a different kind of a network from the first network.

4

26. **(Currently Amended)** The multi-point conference device as defined in claim 19, wherein the image processing unit comprises:

a memory unit for storing an image in accordance with a codec of a speaker terminal as a result of speaker detection by said medium processing unit;

a decoder unit for decoding an image of a speaker held in said memory unit;

a reference image memory unit for holding a reference image obtained on decoding by said decoder unit the last image of a speaker saved in said memory unit; and

an encoder unit for re-encoding an image obtained on decoding by said decoder unit as an image received by a receive unit after a speaker is detected based on a reference image held in said reference image memory unit; wherein

at least the first frame of the image of a speaker received by said receive unit after a speaker is detected is encoded as an intra frame; ~~thereby~~ in a manner enabling handling a case where ~~in which~~ plural items of image data are transmitted by a plurality of terminals connected to a heterogeneous network~~being coped with~~.

27. **(Previously Presented)** A method of performing speaker switching by a multi-point conference device including a medium processing unit for detecting a speaker and an image processing unit for encoding the first image of a speaker received by a receive unit after a speaker is detected as an intra frame, said multi-point conference device switching an image of a speaker by transmitting an intra frame to non-speaker terminals participating in a conference, said method comprising the steps of:

determining whether or not the image of a speaker received is an intra frame;

stopping the processing of said image processing unit and transmitting an intra frame received from a speaker when an intra frame is detected; and

continuing the processing of said image processing unit when it is determined that the image of said speaker is not an intra frame.

28. **(Previously Presented)** A method of performing speaker switching by a multi-point conference device, connected to a plurality of terminals, comprising the steps of:

transmitting an intra frame transmission request to a terminal when said multi-point conference device detects a speaker; and

the terminal receiving an intra frame transmission request from said multi-point conference device and transmitting an intra frame to said multi-point conference device.

29.   (Currently Amended)   The method as defined in claim 27, wherein said multi-point conference device encodes the first image of a speaker received by a receive unit after a speaker is detected as an intra frame, and transmits the intra frame to non-speaker terminals participating in a conference to control switching of speaker images, said method comprising the steps of:

stopping the processing of said image processing unit and transmitting an intra frame of a speaker received by a receive unit when it is ~~detected~~ <u>determined</u> that the image of a speaker received by said receive unit is an intra frame; and

continuing the processing of said image processing unit when it is ~~detected~~ <u>determined</u> that the image of a speaker is not an intra frame;

thereby coping with a case wherein a plurality of codecs for image data are transmitted by plurality of terminals connected to a heterogeneous network.

30.   (Previously Presented)   The method as defined in claim 28, comprising the step of:

detecting by a multi-point conference device, a speaker from a plurality of terminals connected to a heterogeneous network.

31.   (Currently Amended)   The method as defined in claim 27, comprising the steps of:

detecting switching of a speaker by said multi-point conference device connected to a plurality of terminals; and

re-encoding by said multi-point conference device the first image as an intra frame and subsequent frames as inter frames when decoding and re-encoding image data received after a speaker is detected, after said speaker detection, and transmitting the image data to non-speaker

terminals; ~~thereby~~ <u>with</u> said non-speaker terminals being capable of decoding an intra frame at the time of speaker switching.

32.   (Previously Presented)   The multi-point conference device as defined in claim 19, wherein said image processing unit, after the speaker detection, re-encodes the first image as an intra frame and subsequent frames as inter frames when decoding and re-encoding image data received after a speaker is detected, and transmits the image data to non-speaker terminals; wherein

said non-speaker terminals are capable of decoding an intra frame at the time of switching of a speaker.

33.   (Previously Presented)   The multi-point conference device as defined in claim 19, further comprising:

a receive unit for receiving a packet from terminals communicatively connected;

a transmission unit for transmitting a transmission packet;

a call processing unit for performing call processing;

a conference control unit for managing the information of conference participants; and

a memory unit for accumulating image data from terminals participating in a conference corresponding to each terminal;

said image processing unit including a decoder unit, a reference image memory unit, and an encoder unit; wherein

said conference control unit, responsive to a speaker detection result from said medium processing unit, notifies said image processing unit of notification to start processing for speaker switching;

said image processing unit, on receipt of said notification to start processing for speaker switching from said conference control unit, selects the accumulation image data targeted for switching from image data from terminals accumulated in said memory unit to copy the selected image data from said memory unit and the decoder unit decodes the copied image data and accumulates the last image decoded in said reference image memory unit as a reference image;

7

said image processing unit receives the image data targeted for switching from said receive unit, said image data being supplied to said decoder unit when said image data is not an intra frame,

said decoder unit performs decoding processing according to said reference image accumulated in said reference image memory unit, said decoded image data being re-encoded by said encoder unit, the re-encoded image data being supplied to said medium processing unit;

said medium processing unit mixes the re-encoded image data to be transmitted to non-speaker terminals to supply the resulting data to said transmission unit; and wherein

said transmission unit packetizes the image data from said medium processing unit to transmit the packetized data to said terminals.

34.   (Currently Amended)   The multi-point conference device as defined in claim 33, wherein said receive unit checks image data received from a speaker terminal during the time between speaker detection by said medium processing unit and saving of said reference image in said reference image memory unit by said image processing unit; and wherein

when said image data is an intra frame, said receive unit stops supplying said image data to said decoder unit, said image data being supplied to said medium processing unit, ~~thereby~~ with processing for speaker switching being completed.

35.   (Currently Amended)   The method as defined in claim 27, comprising the steps of:

storing image data from a terminal participating in a conference in a memory unit;

detecting a speaker;

decoding an image data of a speaker targeted for switching stored in said memory unit and accumulating the last image decoded in a reference image memory unit as a reference image upon speaker detection;

~~deciding~~ determining whether ~~or not~~ an image data received from a speaker terminal after speaker detection is an intra frame;

decoding the image data based on said reference image accumulated in said reference image memory unit in case of the ~~decision~~ determining result ~~not indicating an~~ indicating no intra frame, re-encoding the decoded image data wherein the first image data from said speaker

terminal is re-encoded at the time of speaker switching as an intra frame in the re-encoding process, transmitting said re-encoded image data to non-speaker terminals participating in a conference; and

transmitting an intra frame received from said speaker terminal to non-speaker terminals participating in a conference in case of the decision result indicating an intra frame.

36. **(Previously Presented)** The method as defined in claim 27, comprising:

a first step of decoding encoded image data received from terminal of a speaker which is targeted for switching at the time of speaker switching; and

a second step of re-encoding said decoded image data; wherein

the first image data from a speaker terminal at the time of speaker switching is encoded as an intra frame in the re-encoding process of said second step; and

an intra frame is transmitted to non-speaker terminals participating in a conference at the time of speaker switching.

37. **(Previously Presented)** The conference system as defined in claim 20, wherein said image processing unit comprises:

decoding means for decoding encoded image data transmitted by

from a terminal of a speaker targeted for switching at the time of speaker switching; and

encoding means for re-encoding said decoded image data; wherein

said encoding means encodes the first image data from a speaker terminal at the time of speaker switching as an intra frame when re-encoding said image data; and

an intra frame is transmitted to non-speaker terminals participating in a conference at the time of speaker switching.